

Reconcile qualitative and quantitative abstractions of metabolism

Application to metabolic model gap-filling

Clémence Frioux

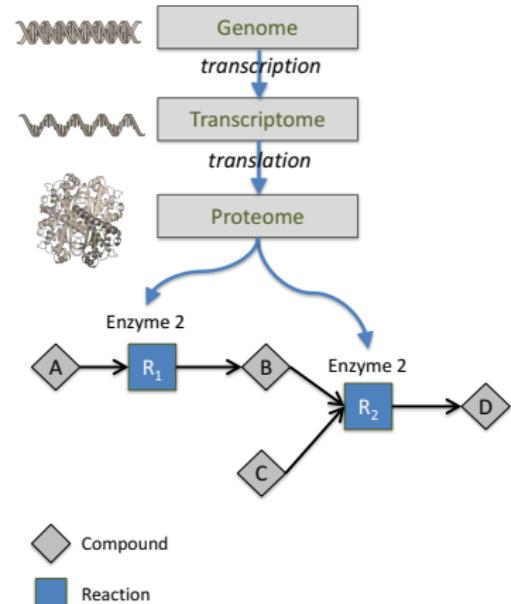
Inria Dyliss - Dynamics, Logics, Inference for biological Systems and Sequences

GT Bioss — March 13, 2017



From biological data to metabolic models

- Integration of omics data
- **Genome-scale data**
- Identification of:
 - ▶ chemical and transport reactions
 - ▶ associated compounds/metabolites
- Acquisition of a network
- Problem: make it **functional**



Identified reactions and compounds form a network that can be further analyzed and completed to reach functionality

Outline

Available data

- Genome-scale metabolic models
- Reactions
- Compounds

What I am interested in

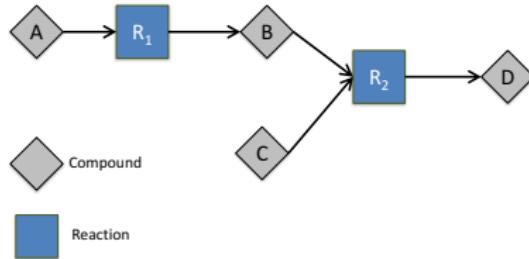
Activation of reactions
↓
Production of compounds of interest

State of the art in modeling

- Two main **formalisms** for exploring models
- Numerical (flux)
- Topology (graph)

Which modeling to describe functionality ?
How to get the most out of the two modelings ?

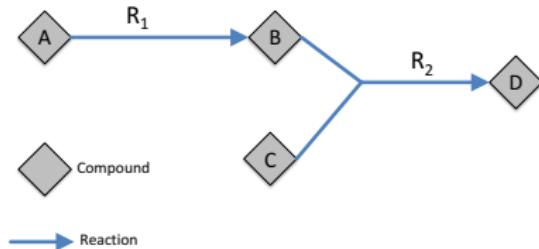
Representation



- Graph topological object G

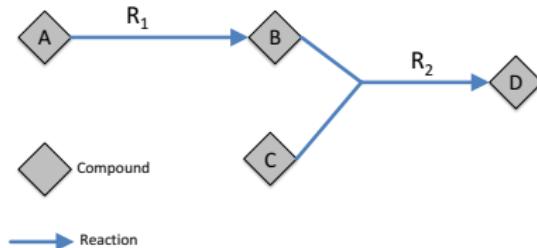
- ▶ bipartite directed graph
$$G = (R \cup M, E)$$

Representation

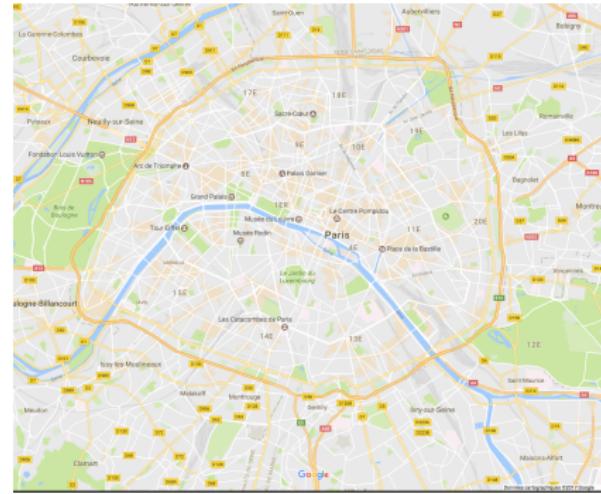


- Graph topological object G
 - directed hypergraph
- $$G = (M, R)$$

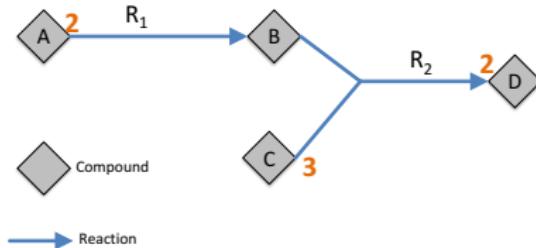
Representation



- Graph topological object G
- directed hypergraph
$$G = (M, R)$$

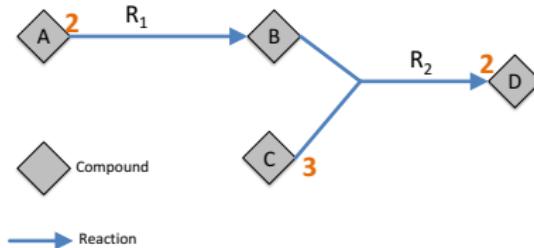


Representation



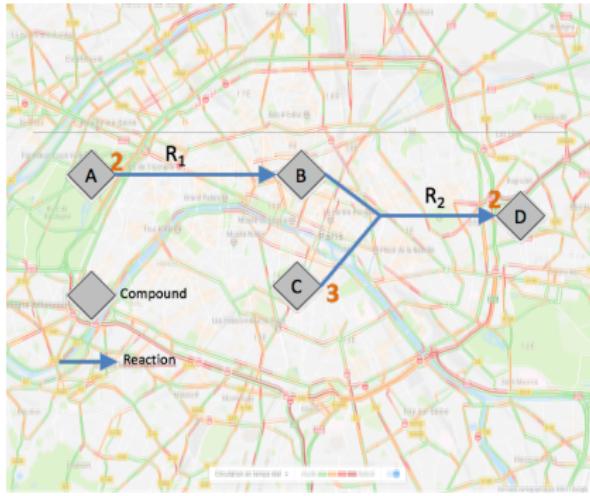
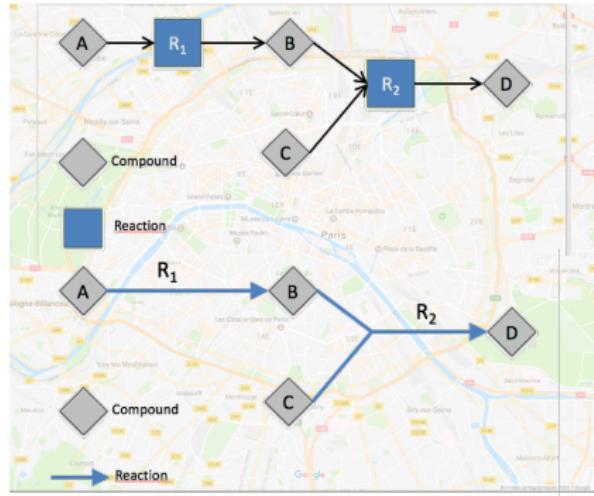
- Addition of numerical values and constraints $\{G, stc, fluxes, bounds\}$
 - ▶ flux v_r of reaction $r =$ bounded numerical value describing the rate of activation of r
 - ▶ stc = stoichiometry, numerical value attributed to every compound of reactions

Representation



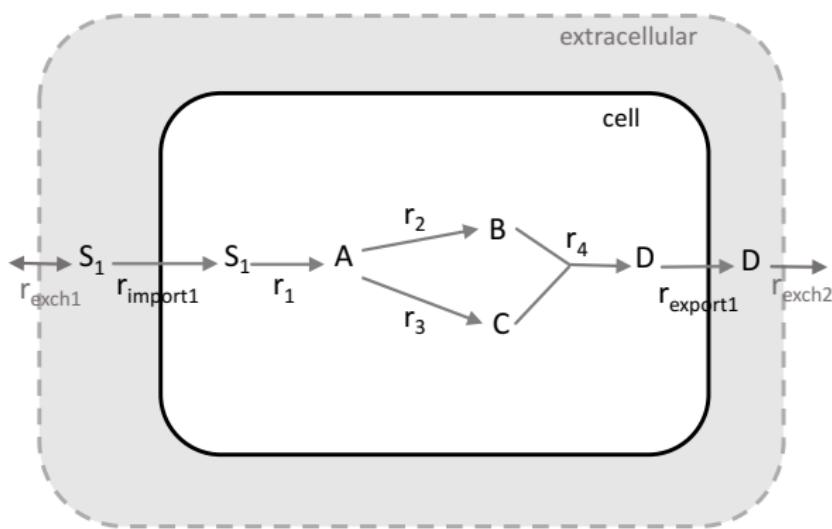
- Addition of numerical values and constraints $\{G, stc, fluxes, bounds\}$
 - ▶ flux v_r of reaction r = bounded numerical value describing the rate of activation of r
 - ▶ stc = stoichiometry, numerical value attributed to every compound of reactions

Representation



2 graph representations, a topological or a numerical point of view

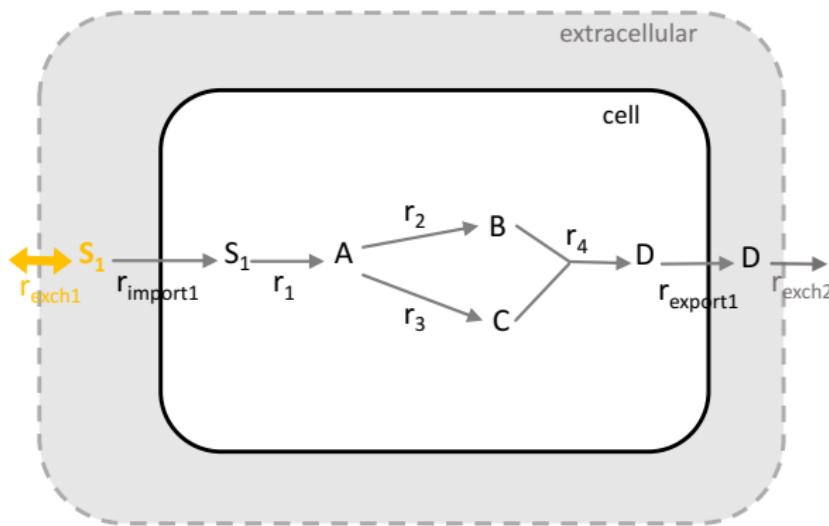
Inputs and objectives of a model



Inputs and objectives of a model

Growth medium — inputs to the model

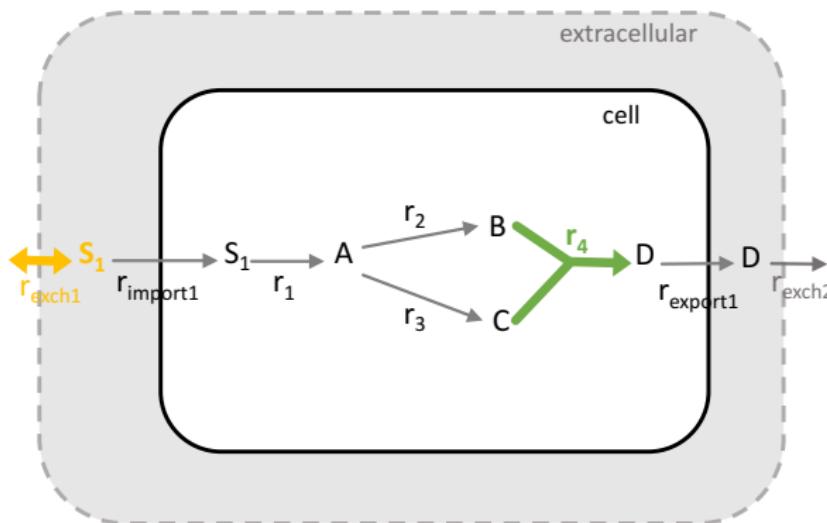
- seeds = attributes of the graph
- products of reactions without reactants



Inputs and objectives of a model

Biomass — Objective reactions of the model

- explicitly defined
- biomass or any other reaction of interest



Objective reactions functionality depends on the inputs of the model

Flux Balance Analysis (FBA) for steady-state accessibility

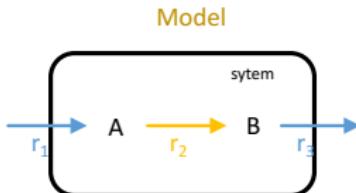
Accessible reactions A_G^F in graph G based on FBA

- To assess accessibility of reaction r
- Maximize v_r with FBA under steady-state assumption $S.v = 0$
- if $\max(v_r) > 0$: $r \in A_G^F$

FBA is the most used method to evaluate quality of metabolic models

Quantitative accessibility of reactions can be achieved with the **FBA** paradigm

Flux Balance Analysis (FBA) for steady-state accessibility



Stoichiometric matrix

$$\begin{matrix} & r_1 & r_2 & r_3 \\ A & 1 & -1 & 0 \\ B & 0 & 1 & -1 \end{matrix} = S$$

Steady-state assumption
compounds concentrations are stable over time

$$\frac{d}{dt} \begin{bmatrix} [A] \\ [B] \end{bmatrix} = 0 \Leftrightarrow S \cdot v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of A per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of B per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

- Fluxes are bounded
 $0 \leq v \leq 2$
- R_2 reaction is maximized
 $\max v_2 = ?$

Optimal flux distribution

?

Adapted from
 Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Flux Balance Analysis (FBA) for steady-state accessibility



Steady-state assumption
compounds concentrations are stable over time

$$\frac{d}{dt} \begin{bmatrix} A \\ B \end{bmatrix} = 0 \Leftrightarrow S \cdot v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of A per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of B per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

- Fluxes are bounded
 $0 \leq v \leq 2$
- R_2 reaction is maximized
 $\max v_2 = ?$

Optimal flux distribution

?

Adapted from
Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Flux Balance Analysis (FBA) for steady-state accessibility



Steady-state assumption

compounds concentrations are stable over time

$$\frac{d}{dt} \begin{bmatrix} A \\ B \end{bmatrix} = 0 \Leftrightarrow S \cdot v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of **A** per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of **B** per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

- Fluxes are bounded
 $0 \leq v \leq 2$
- R_2 reaction is maximized
 $\max v_2 = ?$

Optimal flux distribution

?

Adapted from
Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Flux Balance Analysis (FBA) for steady-state accessibility



Steady-state assumption
compounds concentrations are stable over time

$$\frac{d}{dt} \begin{bmatrix} A \\ B \end{bmatrix} = 0 \Leftrightarrow S \cdot v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of **A** per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of **B** per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

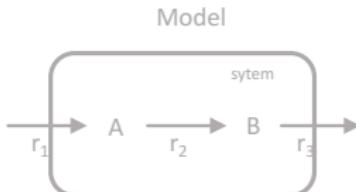
- Fluxes are bounded
 $0 \leq v \leq 2$
- R₂ reaction is maximized
 $\max v_2 = ?$

Optimal flux distribution

?

Adapted from
Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Flux Balance Analysis (FBA) for steady-state accessibility



Stoichiometric matrix

$$\begin{array}{c} r_1 \quad r_2 \quad r_3 \\ \hline A & [1 & -1 & 0] \\ B & [0 & 1 & -1] \end{array} = S$$

Steady-state assumption
compounds concentrations are stable over time

$$\frac{d}{dt} \begin{bmatrix} A \\ B \end{bmatrix} = 0 \Leftrightarrow S \cdot v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of A per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of B per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

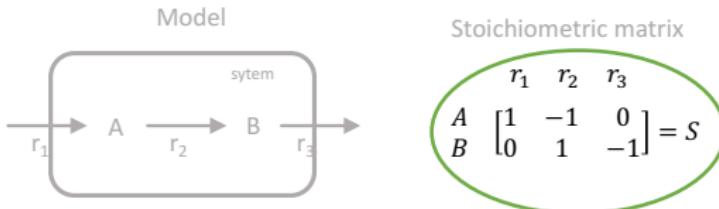
- Fluxes are bounded
 $0 \leq v \leq 2$
- R₂ reaction is maximized
 $\max v_2 = ?$

Optimal flux distribution

?

Adapted from
 Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Flux Balance Analysis (FBA) for steady-state accessibility



Steady-state assumption
compounds concentrations are stable over time

$$\frac{d[A]}{dt} = 0 \Leftrightarrow S.v = 0 \Leftrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$1 \times v_1 - 1 \times v_2 =$ change in conc. of A per unit of time
 $(\text{mol. amount of } A \text{ formed by } r_1 \times \text{rate of } r_1 - A \text{ consumed by } r_2 \times \text{rate of } r_2)$
 $1 \times v_2 - 1 \times v_3 =$ change in conc. of B per unit of time
 $(\text{mol. amount of } B \text{ formed by } r_2 \times \text{rate of } r_2 - B \text{ consumed by } r_3 \times \text{rate of } r_3)$

Additional constraints

- Fluxes are bounded
 $0 \leq v \leq 2$
- R₂ reaction is maximized
 $\max v_2 = 2$

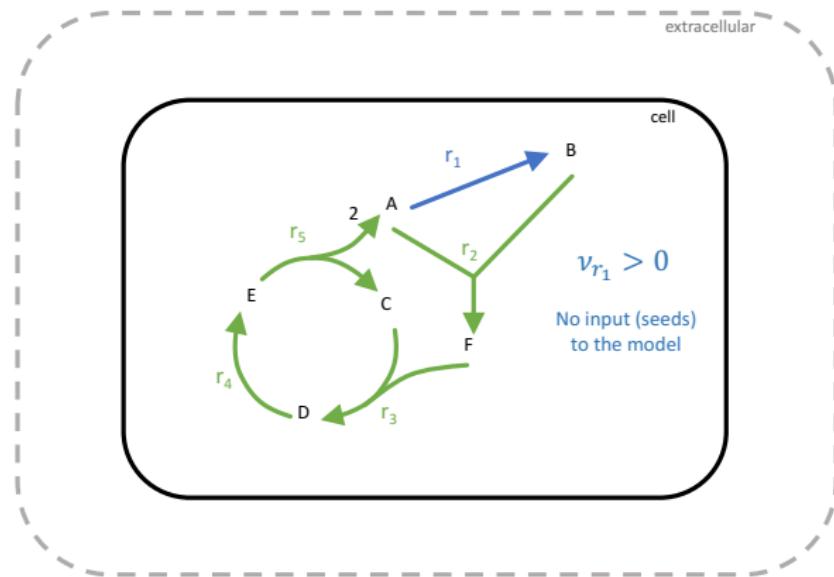
Optimal flux distribution

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

Adapted from
Kim, M. K., & Lun, D. S. (*Biotechnology Journal* 2014).

Limits of FBA-based accessibility

- thermodynamically infeasible cycles
- highly dependant on stoichiometry and thus quality of data



A focus on steady-state might **miss information** about the model

Topological accessibility to retrieve the initial state of the model

Accessibility based on topology

- stoichiometric coefficients are not taken into account
- propagate accessibility of reactions from the seeds
- easy to compute

Qualitative accessibility of reactions can be propagated from the inputs of the graph



Topological accessibility to retrieve the initial state of the model

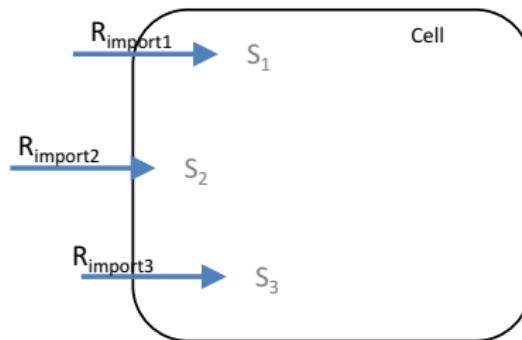
Reachable compounds from seeds S : $\Sigma_G(S)$

For any compound m , $m \in \Sigma_G(S)$ iff:

- $m \in S$
- $m \in \text{products}(r)$ s.t. $\text{reactants}(r) \subseteq \Sigma_G(S)$, r being a reaction

Accessible reactions A_G^T :

for any reaction r , $r \in A_G$ iff $\text{reactants}(r) \subseteq \Sigma_G(S)$



Recursive propagation of topological accessibility from the inputs of the graph

Topological accessibility to retrieve the initial state of the model

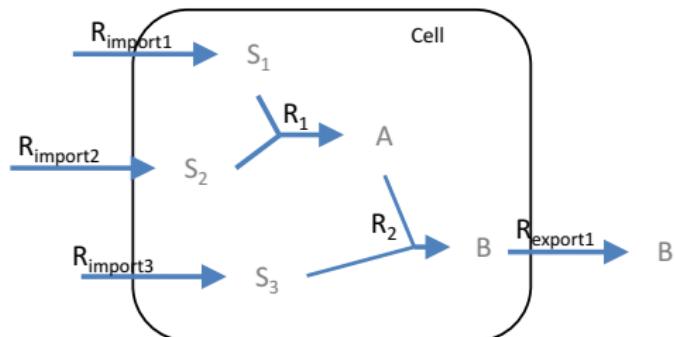
Reachable compounds from seeds S : $\Sigma_G(S)$

For any compound m , $m \in \Sigma_G(S)$ iff:

- $m \in S$
- $m \in \text{products}(r)$ s.t. $\text{reactants}(r) \subseteq \Sigma_G(S)$, r being a reaction

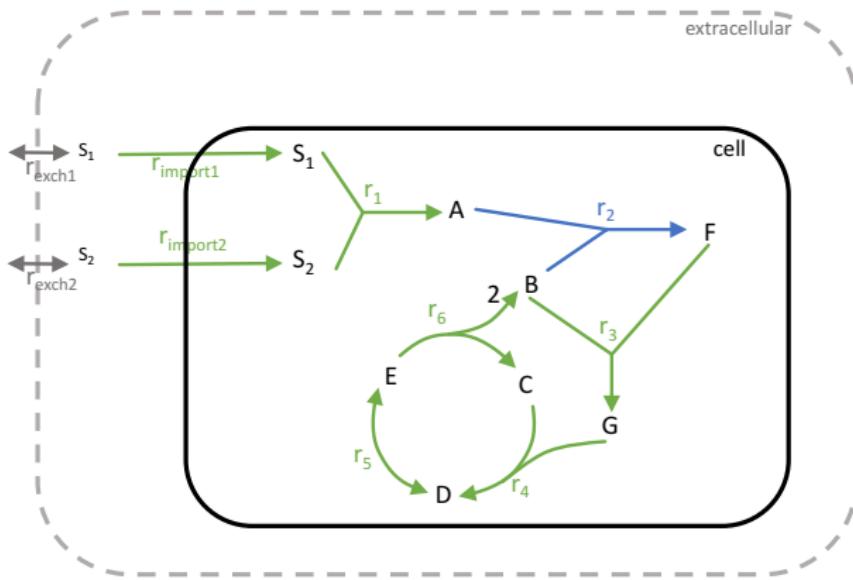
Accessible reactions A_G^T :

for any reaction r , $r \in A_G$ iff $\text{reactants}(r) \subseteq \Sigma_G(S)$



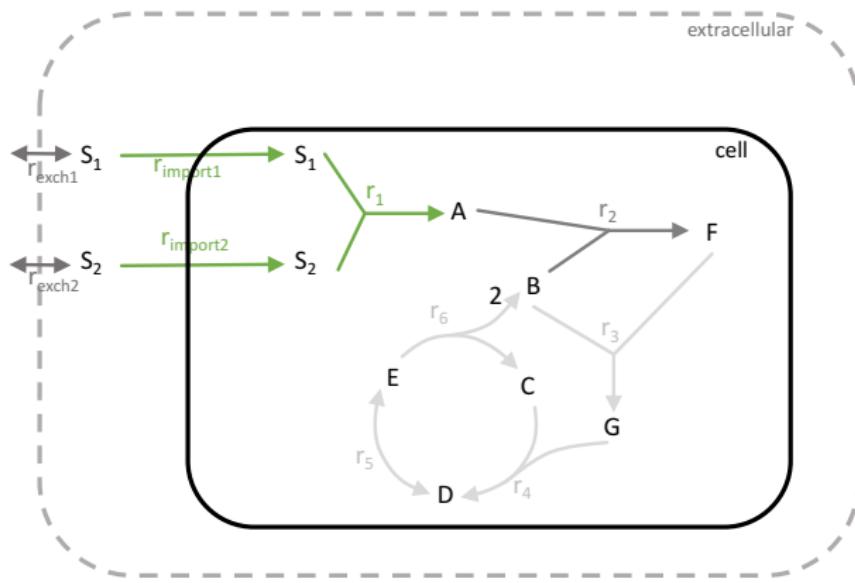
Recursive propagation of topological accessibility from the inputs of the graph

Topological modeling and initial state



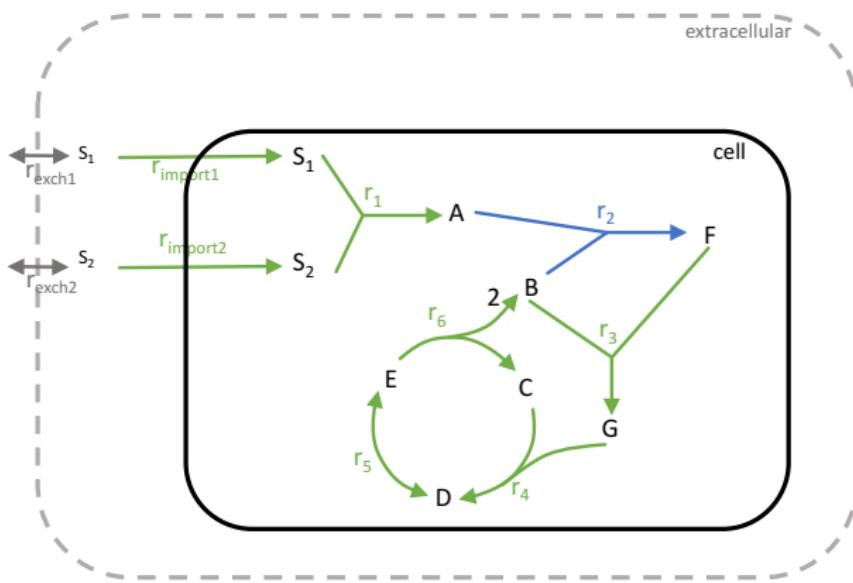
All fluxes are balanced: r_2 is **flux-accessible**

Topological modeling and initial state



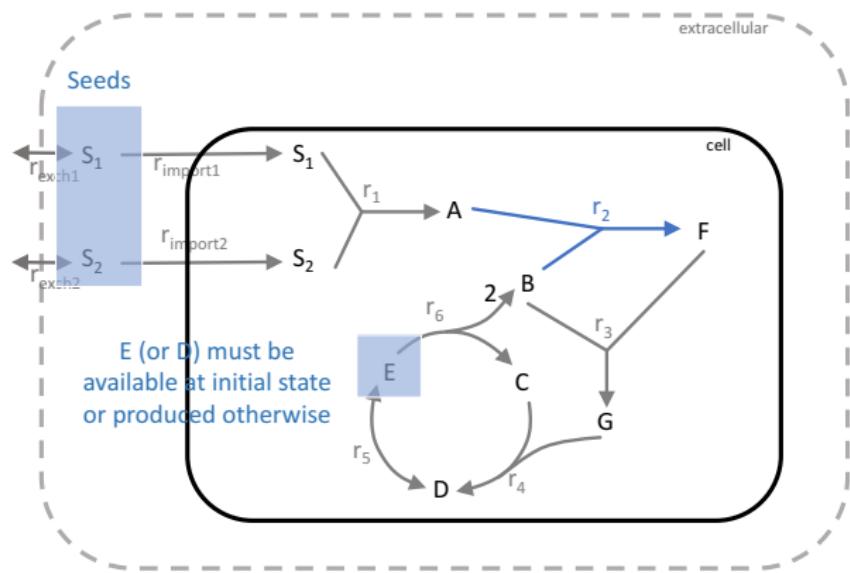
r_2 is **not accessible topologically**. The $\{C, D, E\}$ cycle is not initiated

Topological modeling and initial state



$\{C, D, E\}$ cycle **has to be initiated** prior to steady-state

Topological modeling and initial state



Topological analysis identifies D or E as needed to initiate the cycle
 Topology gives information about the initial state of the system

What to do with these two methods ?

2 techniques to model accessibility of reactions in a metabolic model

- quantitative (steady-state)
 - ▶ metabolites concentrations are constant
 - ▶ no information about their value though
- qualitative (initial-state)
 - ▶ some metabolites concentrations needs to be positive

Can we combine them ? What information would we gain ?

Which added value for the use of **combined flux and topological modelings** ?

Hybrid modeling

- a reaction r of G is hybrid accessible i.e $r \in A_G^H$ iff
 $r \in A_G^F$ and $r \in A_G^T$
- both flux accessible and topologically accessible

Hybrid modeling takes **both initial and steady states** into account to define accessibility of reactions



ASP modulo LP

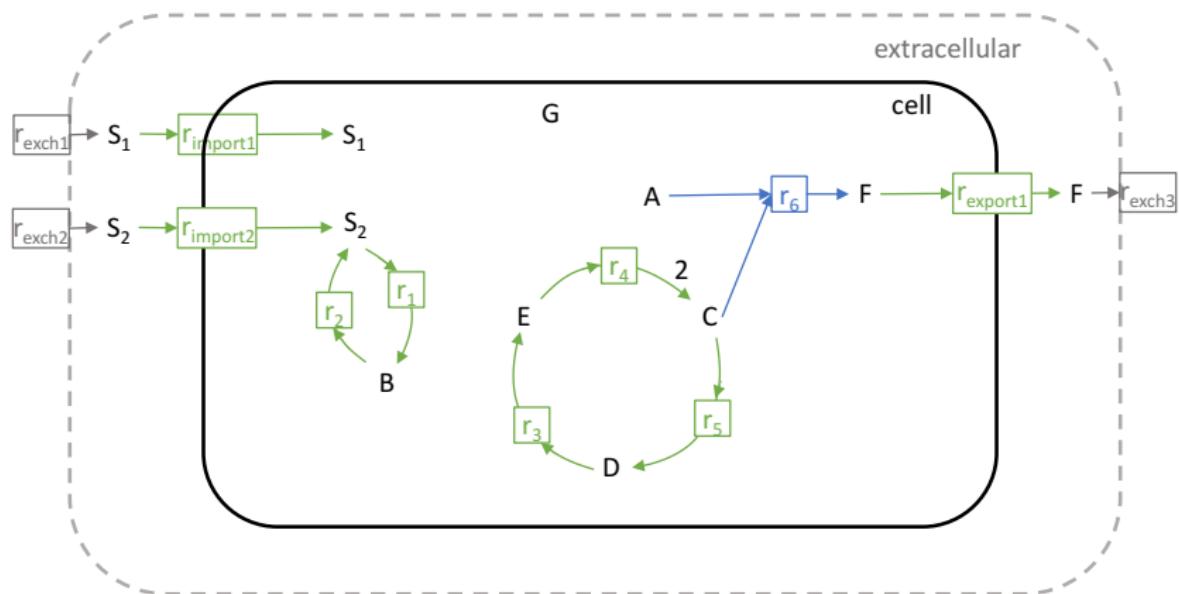
- Answer Set Programming (ASP)
 - ▶ programming paradigm
 - ▶ efficient solving for combinatorial problems (Clingo)
- Linear Programming (LP)
 - ▶ solving of FBA
 - ▶ Cplex or other solvers (Lpsolv)
- *new !* ASP modulo LP
 - ▶ **propagation** of linear constraints
 - ▶ solving of hybrid problems

Application to gap-filling of metabolic models

- Gap-filling
 - ▶ models built from annotation... are not complete
 - ▶ fill "holes" in the model to connect seeds and reactions of interest
- Pluto hybrid gap-filling tool
 - ▶ propose a minimal number of reactions picked from a database
 - ▶ such that reactions of interest
 - ▶ are accessible in flux and topologically

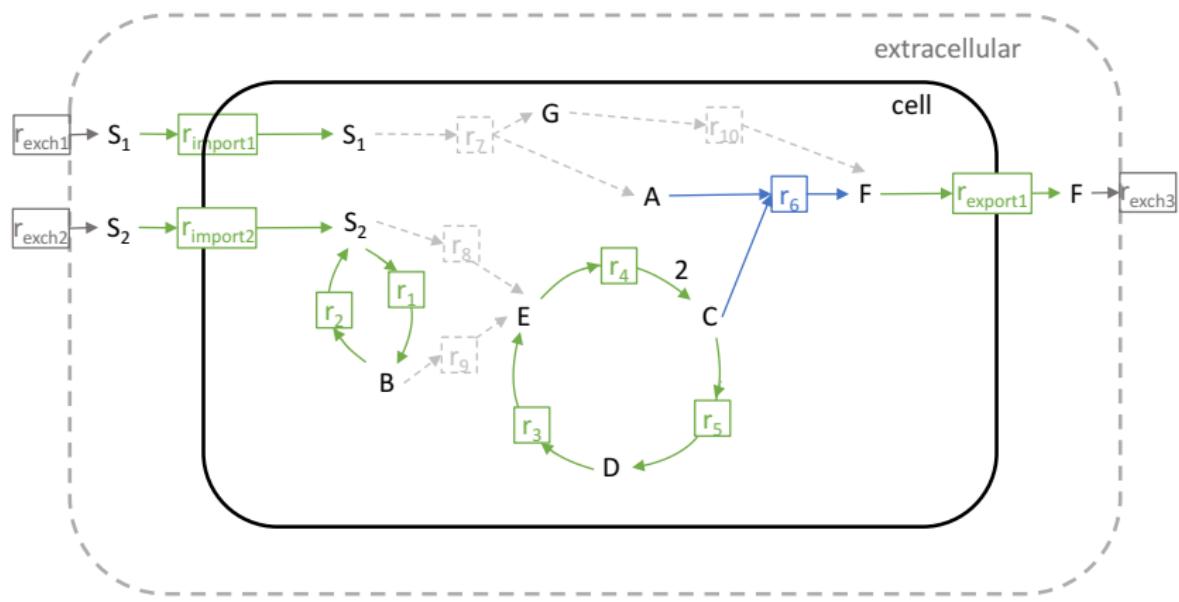
Hybrid gap-filling provides a model that is **functional** at steady-state and takes into account the initial state of the system

Application to gap-filling of metabolic models



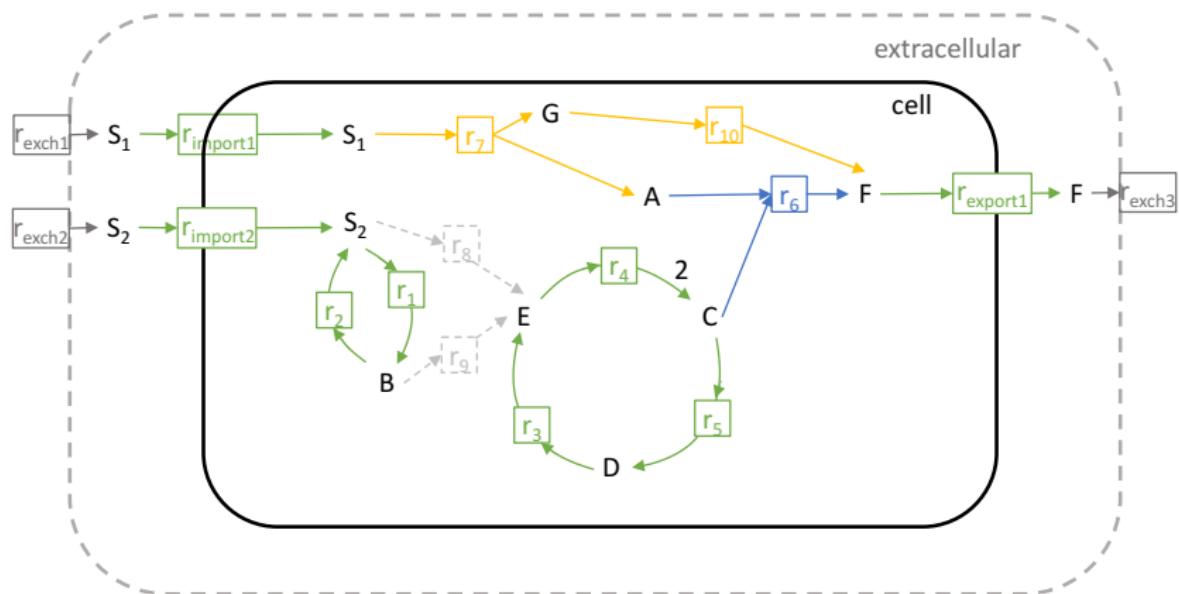
Incomplete model that needs gap-filling to make r_6 accessible

Application to gap-filling of metabolic models



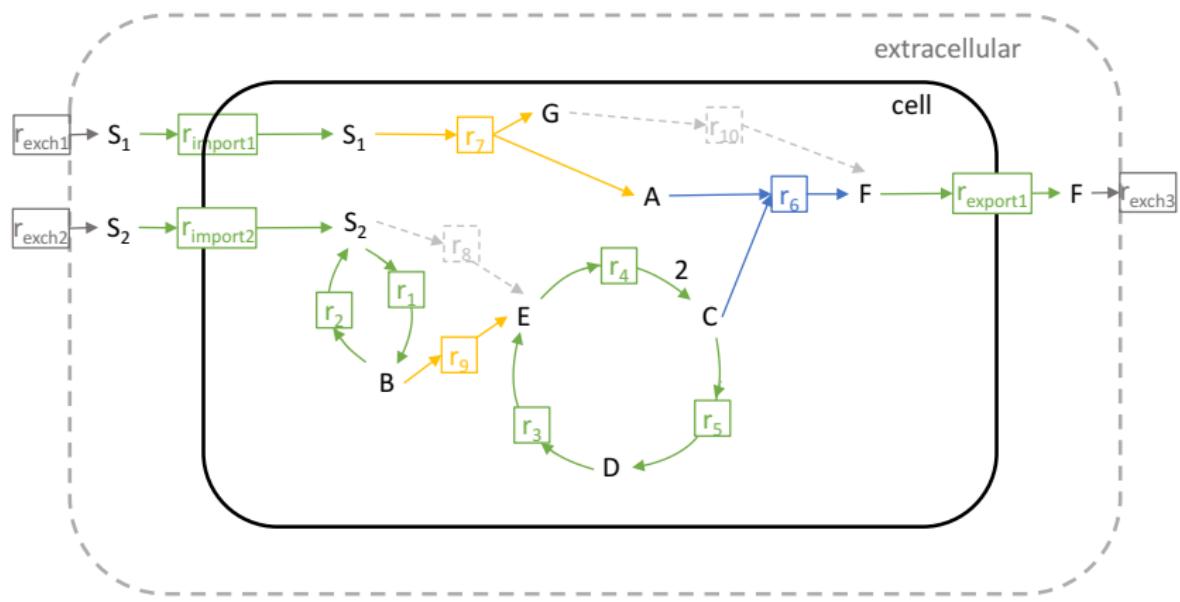
Available reactions for gap-filling

Application to gap-filling of metabolic models



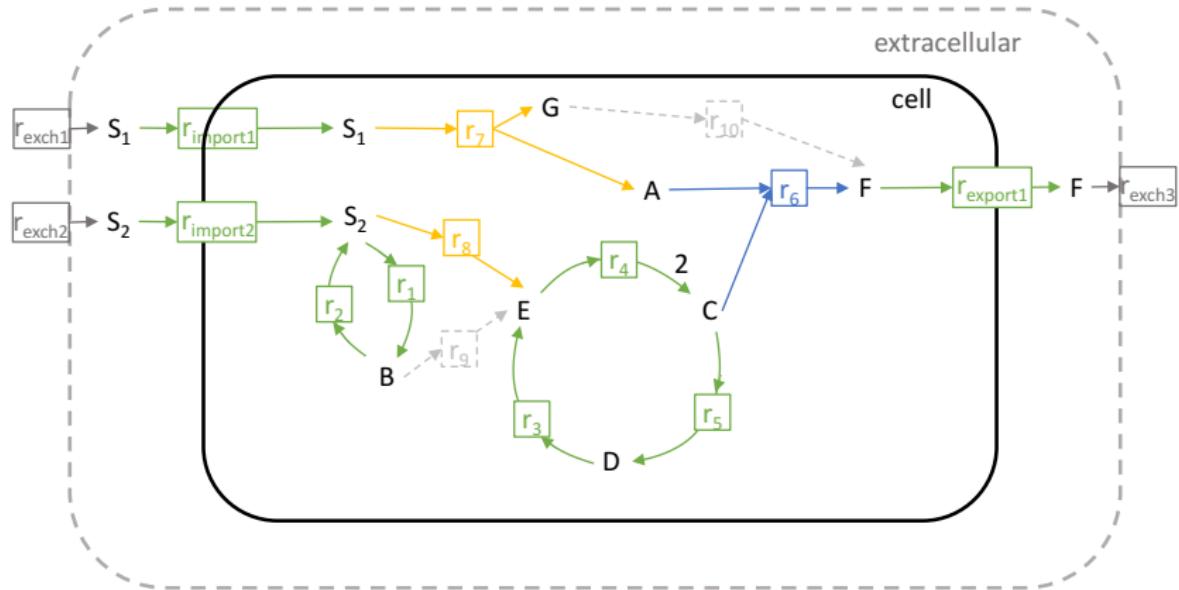
FBA compliant minimal solution (G needs consumption for steady-state balance)

Application to gap-filling of metabolic models



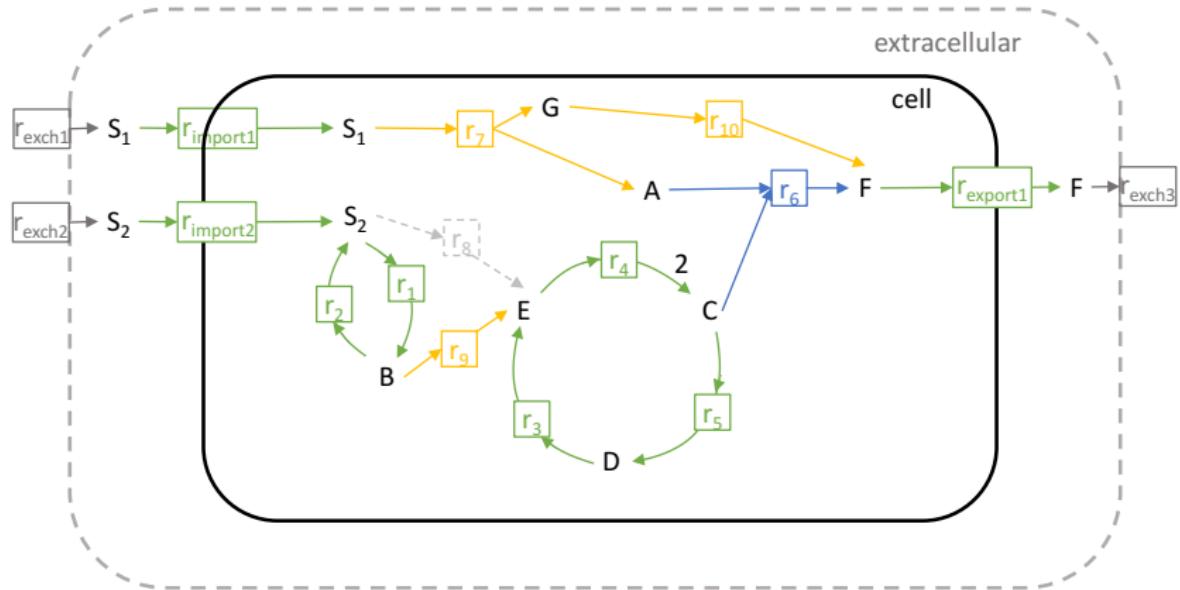
Topological minimal #1 solution: initiates the cycle

Application to gap-filling of metabolic models



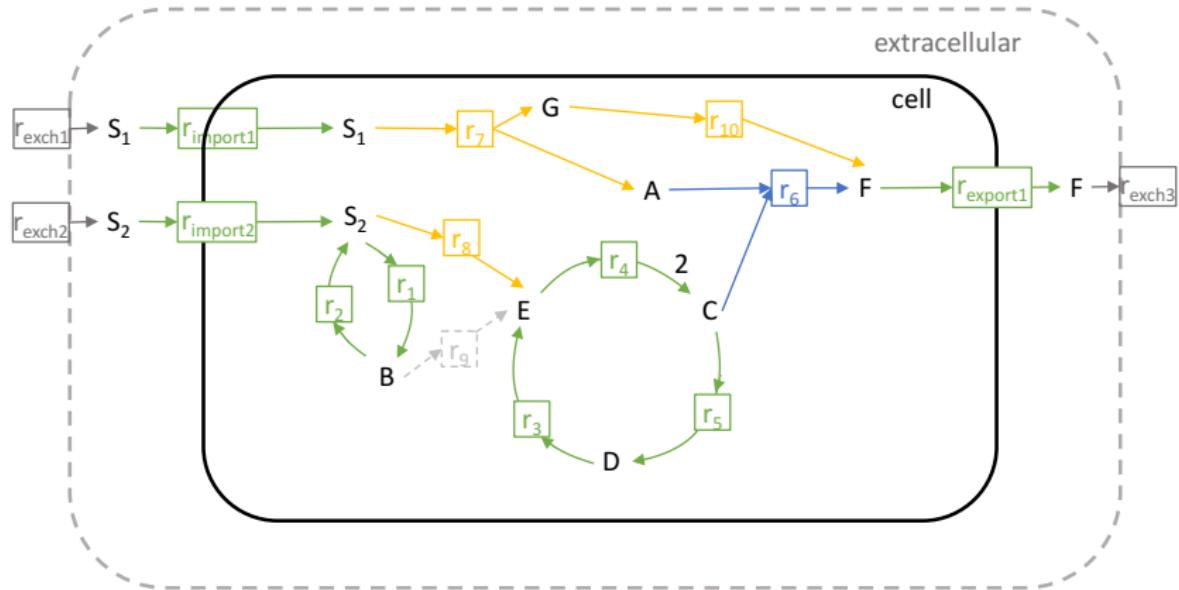
Topological minimal #2 solution: initiates the cycle

Application to gap-filling of metabolic models



Hybrid solution #1: satisfies initial and steady-states

Application to gap-filling of metabolic models



Hybrid solution #2: satisfies initial and steady-states

Results on E. coli degraded models

- Comparison of **biomass flux restoration rates** after **Hybrid**, **Topological** or **Flux** gap-filling
- Percentage of models - **3600 tests** - with restored biomass flux

Degradation
10%
20%
30%
40%
all

Topology restrains search space to access flux-compliant solution more efficiently



Results on E. coli degraded models

- Comparison of **biomass flux restoration rates** after **Hybrid**, **Topological** or **Flux** gap-filling
- Percentage of models - **3600 tests** - with restored biomass flux

Degradation	Topological
10%	73.3%
20%	25.0%
30%	6.8%
40%	3.3%
all	27.1%

Topology restrains search space to access flux-compliant solution more efficiently



Results on E. coli degraded models

- Comparison of **biomass flux restoration rates** after **Hybrid**, **Topological** or **Flux** gap-filling
- Percentage of models - **3600 tests** - with restored biomass flux

Degradation	Topological	Flux (relaxed)
10%	73.3%	6.2%
20%	25.0%	5.7%
30%	6.8%	0.0%
40%	3.3%	0.0%
all	27.1%	3.0%

Topology restrains search space to access flux-compliant solution more efficiently

Results on E. coli degraded models

- Comparison of **biomass flux restoration rates** after **Hybrid**, **Topological** or **Flux** gap-filling
- Percentage of models - **3600 tests** - with restored biomass flux

Degradation	Topological	Flux (relaxed)	Hybrid*
10%	73.3%	6.2%	100%
20%	25.0%	5.7%	96.3%
30%	6.8%	0.0%	86.6%
40%	3.3%	0.0%	75.8%
all	27.1%	3.0%	89.7%

* optimal or suboptimal solutions obtained in 20 min

Topology restrains search space to access flux-compliant solution more efficiently



Conclusion

- hybrid modeling is of high interest in systems biology
- gap-filling is one application among many
- using both quantitative and qualitative criteria might apply to many topics
→ application to ecosystems to fit the sparse available data ?

Credits

Acknowledgments

- Sebastian Schellhorn
- Philipp Wanko
- Anne Siegel & Torsten Schaub

